

University of Massachusetts Medical School

eScholarship@UMMS

University of Massachusetts and New England
Area Librarian e-Science Symposium

2018 e-Science Symposium

Apr 5th, 12:00 AM

Data Rescue

Margaret Janz

University of Pennsylvania

Follow this and additional works at: https://escholarship.umassmed.edu/escience_symposium



Part of the [Scholarly Communication Commons](#)



This work is licensed under a [Creative Commons Attribution 4.0 License](#).

Repository Citation

Janz, M. (2018). Data Rescue. *University of Massachusetts and New England Area Librarian e-Science Symposium*. <https://doi.org/10.13028/bx0h-6a65>. Retrieved from https://escholarship.umassmed.edu/escience_symposium/2018/program/3

Creative Commons License



This work is licensed under a [Creative Commons Attribution 4.0 License](#).

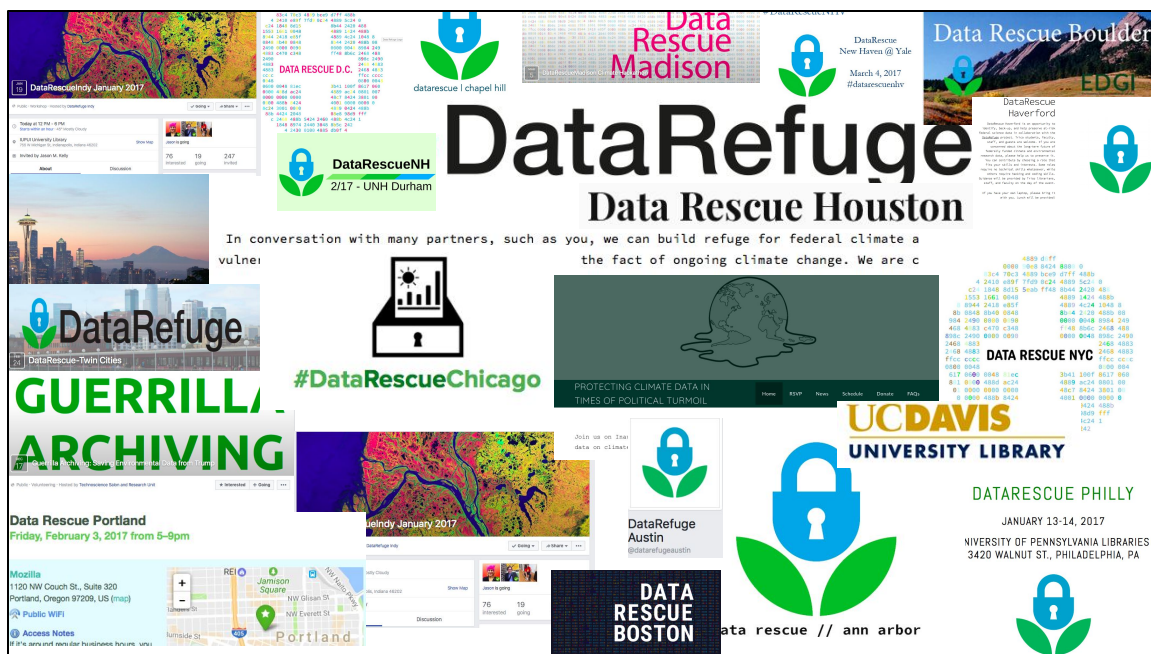
This material is brought to you by eScholarship@UMMS. It has been accepted for inclusion in University of Massachusetts and New England Area Librarian e-Science Symposium by an authorized administrator of eScholarship@UMMS. For more information, please contact Lisa.Palmer@umassmed.edu.

Data Rescue

eScience Symposium Breakout Session
2018-04-05

Photo by [Lukas Juhas](#) on [Unsplash](#)

Intro - my name is Margaret Janz, I am the scholarly communications and data curation librarian at University of Pennsylvania. I'm also a co-founding member of the Data Refuge project (along with Laurie Allen, Bethany Wiggin, and Patricia Kim)



History of Data Refuge in a nutshell: PPEH fellows were worried, we decided to hold a data rescue event, worked with EDGI and others to create workflows for copying data and nominating pages for the End of Term Harvest. Also Storytelling - which I'll talk about more in a hot minute

Around 50 data rescue events that followed some version of our workflow then took place around the country. Great work was done by all who participated and we learned a LOT about "rescuing" data (gonna talk about that term too). And it raised a lot of questions for others.



Endangered Data Week

raising awareness of threats to publicly available data; **exploring** the power dynamics of data creation, sharing, and retention; and **teaching** ways to make endangered data more accessible and secure

[home](#)[about](#)[events & map](#)[past events](#)[news](#)

2018

- [Citizen Science: Health Hackathon](#)
- [Introduction to Tableau](#)
- [Intro to the Unix Shell](#)
- [A Different Kind of Data Rescue: Digitizing & Preserving Obsolete Analog Video](#)
- [Data Sharing in Public Health Emergencies](#)
- [Digital Archives & EDW Trading Cards](#)
- [Endangered Data: Interdisciplinary Panel & Practitioner Lightning Talks](#)
- [Endangered Video Games](#)
- [Intro to Web Scraping](#)
- [Introduction to Data Manipulation with R](#)
- [Introduction to R](#)
- [Public Data Stories](#)
- [ArcGIS Online: Story Maps](#)
- [Basics of Version Control with Git](#)
- [Citizen Analysts: How Regular People Use Public Data to Save the World](#)
- [DataLumos: A Tool for Improving the Future Accessibility of Valuable Government Data](#)
- [#EndangeredData Week Twitter Chat](#)
- [Environmental Justice in Upstate NY](#)
- [Mapping with Tableau](#)
- [Organizing and Using Government Data](#)
- [What Can Be Done to Protect Endangered Government Data?](#)


2017

- [Census Data: Access, Importance, and the Future](#)
- [#critlib Twitter chat: Endangered Data](#)
- [EDW at NDSA: Standards & Practices open call](#)
- [#EndangeredData Twitter chat](#)
- [Indiana University Data Stories](#)
- [Introduction to Libra Data \(Dataverse at UVA\)](#)
- [Personal Digital Archiving](#)
- [What is Endangered Data?](#)
- [Demos & Donuts: Box](#)
- [Endangered Data Brown Bag Discussion](#)
- [Endangered Data Week - Panel](#)
- [EPA archiving work session](#)
- [Introduction to DocNow](#)
- [Introduction to Git/Github](#)
- [Letter Writing Event - Write and Bite Back](#)
- [Web Scraping with R](#)
- [You Can Dig the Same Hole Twice: The Development of a Metadata Scheme for Archaeological Archives](#)
- [An Introduction to Github](#)
- [Archiving Government Data with DataLumos](#)
- [Crash Course in Research Data Management](#)
- [Data Refuge: From a Veteran](#)

endangerreddataweek.org

Endangered Data Week was established by another group, predominately by Brandon Locke and Rachel Mattson. The focus of the week and the related events was on all kinds of data that might be at risk

RESEARCH DATA SHARING WITHOUT BARRIERS
RDA EU
RDA US
CONTACT US
LOGIN
REGISTRATION


ABOUT RDA
GET INVOLVED
GROUPS
RECOMMENDATIONS & OUTPUTS
RDA FOR DISCIPLINES
PLENARIES
EVENTS
NEWS & MEDIA

Data Rescue IG

Home » Working And Interest Groups » Interest Group » Data Rescue IG

IG
Group details


Status: Recognised & Endorsed

Chair (s): Stephen Diggs, David Gallaher, Lesley Wyborn, Denise Hills

Secretariat Liaison: Timea Biro

TAB Liaison: Helen Graves

Case Statement: [Download](#)

 IG Established

Sincere thanks to "Data Rescue IG Emeritus Chair Dr. Elizabeth Griffen*" for her tireless efforts in support of international data rescue work.

The natural sciences possess a rich heritage of data spanning the entire era of research, encompassing both modern electronic formats and analogue ones (paper, film, books, pro-formas, charts, maps, photographic plates) or primitive magnetic tapes. The information in older data is critical for quantifying changes and trends and differentiating between natural or anthropogenic-induced changes. Unfortunately, most historical data have not been converted into electronic datasets. Those data cannot be accessed by present-day research - to the serious detriment of models that predict future changes. Even those that have been 'digitized' (whether catalogues of measurements or the actual observations) are rarely in interoperable, even easily readable, formats. Though essential to research for their unique time-stamp, heritage data are often in deteriorating state, abandoned, or effectively lost (if not actually destroyed). Their profile is unacceptably low; even the RDA does not have an organized effort focussing on this crucial topic.

Since 2010, CODATA has hosted a Task Group for "Data At Risk" (DAR-TG). DAR-TG led an enthusiastic Plenary #3 BoF session, and we are now proposing a new RDA IG for "Data Rescue" with affiliation to

Data Rescue IG

Status: Recognised & Endorsed

Chair(s): Stephen Diggs, David Gallaher, Lesley Wyborn, Denise Hills

Group Email:

Secretariat Liaison: Timea Biro

TAB Liaison: Helen Graves

Public - accessible to all site users

[Join Group](#)

Index	Add new content
Click here to create a wiki index for this group.	
Group Mailing list Archive	

<https://www.rd-alliance.org/groups/data-rescue.html>

The idea of "rescuing" data was also not new when DR started. RDA has had a working group on data rescue since at least 2010. Rescuing data means a great many things to many people - as it should.

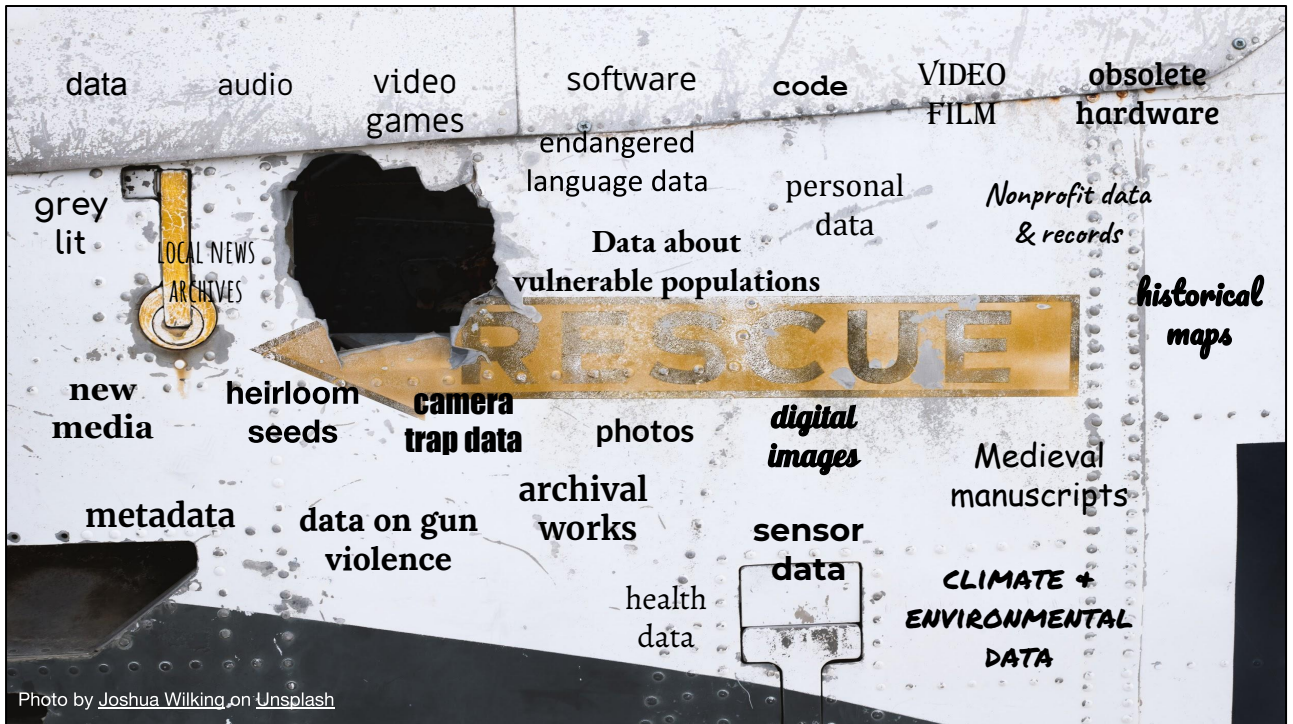


Photo by [Joshua Wilking](#) on [Unsplash](#)

So what might be in need of rescue? Data. Audio. Video games.....literally anything. We rescue what matters to us - as individuals, as institutions, as communities, as ragtag groups of misfits. Figure out what's important to you - some of you probably already have thoughts on this and that's why you came to this session.



The other consideration when considering a data rescue, is why the data need rescuing. There was some misconceptions when DR started that we meant for it to imply that the stewards of govt data were bad at their jobs. They're EXCELLENT at their jobs. Those data are backed up beautifully in most cases. But they're backed up on federal servers, being managed by federal employees. The risk we were concerned with was about access.

The access to these data was at risk because of political factors, which is just one way data can be at risk. That applies to nongovt data, too. Org politics and leadership priorities can put data at risk in the same way.

Technical risks include file formats, software and hardware obsolescence, bit rot, regular rot of physical items, items made from unusual materials. A lack of documentation can also put data at risk; if no one know the data exists or has not context to make sense of the data it is hard to protect it.

How the data is used, and who it's used by can also make data vulnerable - as well as provide protections. The Data Refuge Stories project helps show broad usefulness of data by connecting it to everyday life and human stories.

Legal protections may exist for some data - like retention laws or policies. How or if those policies are enforced can be also put data in risky situations or protect them.

There may also be policies about data deletion, but those policies might not take into account all the users of those data.



Collaboration and partnerships are also key to data rescue. These are new ideas that need new workflows with a variety of stakeholders. We need new solutions that take each others' expertise and needs into account. We have to find new ways to work together. It's not easy to imagine, but I'm going to ask you to do that now.

Describe the Data	Who Uses the Data?
<p>What are they? What formats? How much is there?</p>	<p>Who uses the data? What is the intended use? How else are they used?</p>
What Are the Risks?	Rescue Recommendations
<p>Is the format sustainable? How is it currently cared for? Who cares for it? ...</p>	<p>Partners? Funders? Digitization? Conservation? Citizen Science? Metadata? New home? ...</p>

Attendees of this session were asked to think about these questions and fill out the worksheet provided (with the same questions) with some data they are concerned about in mind. Then we shared some thoughts about rescuing data and help each other think of ways to go about rescuing the data in ways the writers had perhaps not considered. We can help each other so much!